

# ■ Treatment Integrity in a Home-Based Pre-Reading Intervention Programme<sup>†</sup>

Sandra G. van Otterloo\*, Aryan van der Leij and Esther Veldkamp

*Department of Education, University of Amsterdam, The Netherlands*

Treatment integrity is an underexposed issue in the phonological awareness intervention research. The current study assessed the integrity of treatment of the families ( $N = 32$ ) participating in the experimental condition of a home-based pre-reading intervention study. The participating kindergartners were all genetically at risk for developing dyslexia. Two aspects of treatment integrity, the number of lessons completed (quantity) and the quality of the administration of the programme (observed in a videotaped session), were investigated. The level of treatment integrity turned out to be 66% when completion of all lessons was taken as quantitative criterion, and about 74% when quality of the parent-child interaction was assessed. The two measures could predict the pre-reading skills at the end of kindergarten. Together they accounted for 43% of the variance in this dependent variable. Together with pre-reading scores at the pre-test the total predicted variance was 87%. The number of lessons completed still contributed 12% to the prediction after controlling for pre-test scores. The results indicated that treatment integrity indeed appears to be an important aspect of treatment outcome and should therefore be included in intervention studies. Copyright © 2006 John Wiley & Sons, Ltd.

**Keywords:** early intervention; familial at risk; home-based intervention; treatment fidelity; treatment integrity

\*Correspondence to: Sandra G. van Otterloo, Department of Education, University of Amsterdam, PO Box 94208, 1090GE Amsterdam, The Netherlands. Tel.: +31205251359; fax: +31205251200; e-mail: S.G.vanOtterloo@uva.nl

<sup>†</sup>The present study is associated with the research programme 'Early diagnosis and treatment of developmental dyslexia: The role of learning mechanisms, speech manipulation, phonological skills and rapid naming', partly granted by the Netherlands Organization for Scientific Research (NWO). The research programme is a cooperation between the University of Groningen, the Radboud University (Nijmegen), and the University of Amsterdam.

## INTRODUCTION

Troia (1999) wrote a critical review of the experimental methodology of the phonological awareness intervention research. The author evaluated 39 studies on internal and external validity criteria. One of the five most serious methodological shortcomings observed in many studies was insufficient or nonexistent assurance of what he calls treatment fidelity. The other four were: nonrandom assignment of subjects to conditions (54% of studies met criterion), no control for Hawthorne effects (51% of studies met criterion), poor measurement sensitivity (36% met criterion), and inadequately described samples (several criteria were investigated). Troia speaks of treatment fidelity when 'a procedure was used to ensure that treatment conditions were being implemented faithfully' (Troia, 1999, p. 33). As a concept, treatment fidelity is closely related to the concept of treatment integrity, defined by Gresham, MacMillan, Beebe-Frankenberger, and Bocian (2000, p. 198) as 'the degree to which a treatment is implemented as planned or intended'. Because it is more common, treatment integrity will be used throughout the present paper. According to Troia (1999), only five of the 39 investigated studies (13%) met this criterion. Only two studies (5%) also assessed the level of treatment integrity. Most other internal validity aspects were met in 28–92% of the studies. The only internal validity criterion that was met in fewer studies than the assurance of treatment integrity was the 'reporting of effect sizes' (met in 10% of the studies). However, Troia gave this criterion a lower weighting factor than the assurance of treatment integrity. So, treatment integrity is one of the most ignored aspects of internal validity in the phonological awareness intervention research. This ignorance seems unjustified, because when treatment integrity is insufficiently checked and positive training effects are found, the outcome can be caused by something else than the training. In addition, when the training did not have significant results, this could be due to an ineffective treatment as well as an inadequately implemented treatment (Moncher & Prinz, 1991). The present paper aims to demonstrate that treatment integrity can be assured with relatively simple means. Moreover, the results indicate that at least some aspects of treatment integrity affect treatment outcome.

Studies can deal with treatment integrity in several ways. Researchers can promote and stimulate treatment integrity before treatment has started. In addition, they can globally check, verify or monitor the treatment adherence during the course of treatment. Ultimately, they can assess (and report) the level of treatment adherence. Promoting, checking and assessing of treatment integrity can be operationalized differently. Of the reviewed studies that did not ignore treatment integrity, Hatcher, Hulme, and Ellis (1994) trained all treatment agents (teachers) for 3 days in how to use the teaching materials (promotion). The adherence to the protocol was checked by regular meetings and written records for each treatment session. In the study of O'Connor, Jenkins, Leicester, and Slocum (1993) the instructors practiced the treatment format with the investigator every week. All agents met daily to discuss difficulties with application of the treatment. The treatment integrity was checked by regular observations and random audiotaped sessions. In a later study (O'Connor, Jenkins, & Slocum, 1995) the treatment agents (teachers) learned routine formats during two 2-h sessions. During the treatment period new formats were practiced every week in

teacher meetings. All teachers were observed regularly to ensure treatment integrity. The lessons were held as intended in 90–100% of the sessions. In another study of the same authors (Slocum, O'Connor, & Jenkins, 1993) the instructors (graduate students in education) were trained until they reached 100% accuracy in the instructional procedures with children who were not included in the study. Teachers participating in the intervention study of Williams (1980) were asked to fill in log forms that provided, among other things, information about which lesson parts were carried out and the time spent on the programme. These forms were collected during a weekly visit to each classroom. Once every 2–3 weeks this visit also included the observation of one lesson. Like O'Connor *et al.* (1995) Williams reported the percentage of accurate implementations of the treatment. Provided worksheets were used in 70–77% of the time, games were used in 77%, and 91% of the unit pre- and post-tests were used.

In sum, these studies promoted the treatment integrity by (repeated) training of treatment agents. In addition, they checked treatment integrity by meetings, observations, audiotaped sessions and log forms. Only two studies (O'Connor *et al.*, 1995; Williams, 1980) also assessed and reported the level of treatment integrity. However, they did not report precisely how they measured treatment integrity.

There are no other studies, as far as we know, on treatment integrity in the phonological awareness intervention research. There are some reviews, however, of treatment integrity in related domains. We will describe three different reviews on treatment integrity. Table 1 provides an overview of these reviews.

In a review of learning disabilities research Gresham *et al.* (2000) analysed all intervention articles published in the Journal of Learning Disabilities, the Learning Disability Quarterly and in Learning Disabilities Research & Practice from January 1995 to August 1999. In only 12 articles (18%) treatment integrity was assessed and reported. In about half of the articles the treatment integrity was promoted and/or checked, but the articles did not report the level of integrity. Almost one-third of the reviewed articles ignored treatment integrity.

In an earlier review of 181 school-based behavioural intervention studies from 1980 to 1990 (Gresham, Gansle, Noell, Cohen, and Rosenblum, 1993), the authors found that 15% of the reviewed studies assessed and reported levels of treatment integrity. In some studies (10%) the researchers checked the treatment integrity but provided no data on the level of integrity. The other 75% of the studies did not investigate treatment integrity. Gresham *et al.* reported that the mean level of treatment integrity reported in 15% of the studies was 96% (*S.D.* = 5.5%).

Moncher and Prinz (1991) reviewed 359 treatment studies in the domains of clinical psychology, behaviour therapy, psychiatry, and marital and family therapy. They distinguished two issues in treatment integrity, namely treatment integrity in the common way and treatment differentiation. Treatment integrity in the common way refers to 'the degree to which a treatment condition is implemented as intended' (Moncher & Prinz, 1991, p. 247). This concept corresponds to the concept of treatment fidelity as used by Troia (1999) and the concept of treatment integrity of Gresham *et al.* (1993, 2000). Treatment differentiation refers to 'whether treatment conditions differ from one another in the intended manner such that the manipulation of the independent variable actually occurred as planned' (Moncher & Prinz, 1991, p. 248). The majority (55%) of the reviewed studies paid no attention to treatment integrity, about 27% only promoted treatment integrity by the use of a manual and/or the supervision of

Table 1. Overview of mentioning and measuring of treatment integrity in reviews on intervention research

Authors	Year	Domain	Used concept	N Articles reviewed	Promoted and/or checked (%)	Assessed (%)	No information (%)
Troia	1999	Phonological awareness interventions	Fidelity	39	8 <sup>a</sup>	5	87
Gresham <i>et al.</i>	2000	Learning disabilities interventions	Integrity	65	49	18	32
Gresham <i>et al.</i>	1993	School-based behavioural interventions	Integrity	181	10	15	75
Moncher and Prinz	1991	Clinical psychology, behaviour therapy, psychiatry, and marital and family therapy	Integrity	359	27 <sup>a</sup>	18 <sup>b</sup>	55

<sup>a</sup>Both Moncher and Prinz and Troia were less strict than Gresham *et al.* (1993, 2000). They included studies that only promoted integrity, whereas Gresham *et al.* only included studies that at least globally checked integrity.

<sup>b</sup>Moncher and Prinz were less strict than the other researchers, they included all studies that checked the integrity, and not only the one's that also reported the level of integrity.

treatment agents, and 18% checked adherence to protocol. In some studies more methods were used simultaneously. Both the use of supervision and the adherence check increased in time (period 1980–1982 to 1986–1988). In the 67 studies that checked the adherence to the protocol, the information was provided by independent observers only (64%), provided by participants only (21%), provided by both (13%), or was not reported (2%). Gresham *et al.* (1993) argued that the different outcome between their study and the study of Moncher and Prinz (1991) may be partly because of the more stringent criteria used in their review. It seemed that Gresham *et al.* only included studies that globally checked or more precisely measured treatment integrity, whereas Moncher and Prinz (as well as Troia, 1999) also included studies that only promoted treatment integrity.

To conclude, reviews focusing on treatment integrity showed that a subset of intervention studies, ranging from 25 to 67% depending on the field of research and of the criteria used, investigated the treatment integrity of the independent variables in some way. Ten to 49% of the studies only paid attention to promotion and/or checking. Fewer studies also reported the assessed level of treatment integrity (15–18%). In the review of Troia (1999) of phonological awareness interventions, these proportions were even smaller (13 and 5%).

Gresham *et al.* (1993) also computed the correlation between the degree of treatment integrity and the degree of treatment outcome (effect size Cohen's *d*) reported by the reviewed articles. The degree of treatment integrity was positively correlated with the outcome of treatment ( $r = 0.51, p < 0.05$ ). Although correlation is a necessary but not a sufficient condition for causality, the level of treatment integrity probably influences the power of intervention studies. A treatment that is adequately and faithfully implemented creates less error and is therefore a more pure intervention. Moncher and Prinz (1991) also argue that treatment integrity can have an impact on the power of treatment studies, because it reduces unintended variability in the treatment.

## Home-Based Intervention

With regard to pre-reading interventions, most studies, including those evaluated by Troia (1999), focus on intervention at the school of the children. In the present study, we used an adjusted version of the home-based pre-reading programme from our intervention study titled 'Early home-based intervention in the Netherlands for children with genetic risk of reading disabilities', submitted for publication. It may be argued that in home-based interventions, the issue of treatment integrity may be even more important than in school-based phoneme awareness intervention studies. This is because, the treatment agents were nonprofessionals (parents) and they worked with the programme at home instead of in a more neutral environment such as the school. This situation will more easily lead to undesirable variability in the treatment.

Although the use of home-based intervention programmes in this field receives growing interest of researchers (see for example Rack, 2004), only a few studies have been published. Fielding-Barnsley and Purdie (2003) used a programme that was made for children with a family history of reading disability in the year prior to formal schooling. It was based on dialogic reading (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Whitehurst *et al.*, 1994), but with additional instruction focused on rhyme awareness, alliteration, print familiarity and

alphabet knowledge. Prior to the start of the intervention a research assistant presented a videotaped instruction in the family home. Each family also received four pages of written information and a parents' handbook (Barrs & Ellis, 1998). The average number of readings per book turned out to be higher than required, indicating that the parents participated well. The children were tested twice during the first year of formal schooling and outperformed the controls on several reading and spelling measures. The findings indicate that early home-based intervention programmes for at-risk children can be effective. However, although treatment integrity was promoted and checked, the level of treatment integrity was not reported.

Our intervention study, submitted for publication reports the effectiveness of the home-based intervention programme that is also used in the present study. In this study, the authors promoted and checked the treatment integrity by an instruction meeting, regular contact by phone and daily log sheets to be filled in by the parents. The experimental children outperformed the controls on several pre-reading measures right after training. However, the authors did not assess the level of integrity of the parents. So, the possibility cannot be ruled out that the results are influenced or even caused by a variable that was not part of the intervention. The present study that is a part of a second evaluation of the effectiveness of the home-based intervention programme aims to shed light on this question.

### Research Questions and Expectations

The available literature showed that treatment integrity in the phonological awareness research is an underexposed issue. This is unjustified, because treatment integrity influences the validity and the power of intervention studies. The level of treatment integrity even turned out to be strongly correlated with the effect size of intervention studies (Gresham *et al.*, 1993). Because higher treatment integrity in studies seemed associated with larger effect sizes, it is conceivable that within a single study higher treatment integrity of treatment agents is associated with higher scores on the outcome variable. Paradoxically, it may be argued that a higher level of treatment integrity across all participants is related to a smaller variance in that level and, consequently, to a lower correlation with the outcome variable. In other words, a correlation between treatment integrity and the dependent variable is only possible when the variation of treatment integrity across treatment agents is sufficient. In the present study, two research questions were addressed to: (a) How well, in terms of treatment integrity, is the intervention implemented by the tutors? (b) To what extent does treatment integrity contribute to the pre-reading skills at the end of kindergarten?

Although the participation in a daily intervention programme for 14 weeks was quite demanding, we expected that most parents would be motivated to complete the intervention because they were familiar with the consequences of reading problems. However, parents are no professionals, and the quality of the implementation of the programme could therefore vary and even be insufficient in some cases. With regard to the first question, we expected that, in general, the parents would implement the programme quite well. In our intervention study, submitted for publication, most parents seemed to be able to work with the

intervention programme. This conclusion, however, was only based on self-report of the participants.

To understand the context of the present study, it is important to note that the participants in the training made more progress on some pre-reading skills during the final month of kindergarten compared to a no-training control group (paper is in the phase of preparation; in general, the results replicated the findings of our submitted intervention study). We therefore concluded that the training positively influenced the pre-reading skills of the participants. The first question investigated whether this conclusion was supported by the fact that the parents applied the programme with sufficient integrity. With regard to the second question, we expected that treatment integrity in the current study would be positively correlated to pre-reading skills at the end of kindergarten and would contribute to the prediction of pre-reading skills at the end of kindergarten.

### **Some Remarks About the Dutch Educational System**

To be able to interpret the findings, the Dutch educational system and tradition may need some clarification. In the Dutch educational system formal reading starts in the school year when the children reach the age of seven, a year later than in the United Kingdom and related countries but a year earlier than in Scandinavian countries. In the pre-reading phase, there is some attention paid in kindergarten to emergent literacy skills (for example, story telling and reading to the children), but letter knowledge and phonological abilities are not stimulated in a systematic or intensive way. The same is true for the parents at home. Reading to the children is very common but there is no tradition in learning the alphabet at home or stimulation of specific sub-skills of reading. However, children gather letter knowledge incidentally, e.g. by imitating older children at home or watching television ('Sesame street'). Most children can write their own name and, possibly, other familiar names when they enter grade one and may recognize quite a few letters. Instruction and practice in grade one is focussed on the letter-sound connections in the context of short, familiar words, and very intensive (6–7 h a week). As in most transparent orthographies, much attention is paid to phonics and letter-by-letter reading, followed by blending into words. Dutch qualifies as a relatively easy-to-learn orthography, comparable to German and most Scandinavian languages (except Danish that has a more inconsistent orthography). While Finnish, Spanish and Italian orthography appear to be even more transparent, French, Danish and, in particular, English are more difficult to learn than Dutch (Seymour, Aro, & Erskine, 2003). At the end of grade one, most Dutch children are able to read both simple familiar and unfamiliar (non) words with accuracy above 80%.

## **METHOD**

### **Sample**

Thirty-seven children were selected for participation in the experimental group. One girl was excluded because she already could read at pre-test. Four families

Table 2. Control measures and background measures ( $N = 32$ )

Measure	<i>M</i>	<i>S.D.</i>
Age in month (Sept. 1st, kindergarten 2)	64.71	3.26
Receptive vocabulary <sup>a</sup>	72.35	9.64
Nonverbal IQ <sup>b</sup> (Raven's CPM)	25.12	4.89
Raven percentile scores <sup>b</sup>	72.63	25.35
Measure	University or college degree (%)	Other (%)
Education level of the mother	53	47
Education level of the father	44	56

<sup>a</sup>Measured at pre-test.<sup>b</sup>Measured halfway first grade.

drew back at an early stage: two families because the parents turned out to be too occupied to participate, one family because the intervention was too difficult for the parents to administrate and in one family the parents did not agree with the experiment. Participants were 32 children (20 boys, 12 girls) in their second kindergarten year. The children were derived from 25 schools. Per school a mean of 1.28 children (range 1–3) participated in the experiment. The children were genetically at risk for developing dyslexia. To select them we used criteria comparable to similar studies of at-risk children in the pre-reading phase (e.g. Elbro & Petersen, 2004; Hindson *et al.*, 2005). Parents, who had experienced dyslexia, were invited to join the study. At least one had to meet the selection criteria indicating that their performance on standardized word recognition and phonological decoding tests was far below expectation. (More details will be presented in our submitted intervention study).

To describe the selected children, Table 2 shows the age and IQ of the children, and their background (education level of the parents). Compared to a norm (Verhoeven & Vermeer, 1996), the mean score of the sample on receptive vocabulary was high. The mean score on the Raven corresponds with a percentile score of 73 (van Bon, 1986). However, this norm is quite old, as a result of which the scores might be a bit overestimated (Flynn-effect). Probably due to the voluntary basis of the selection procedure the children possessed above-average receptive vocabulary and nonverbal IQ. Moreover, the sample of parents was relatively highly educated.

## Procedure

All subjects participated in a home-based pre-reading tutor programme, which was carried out in the second half of the final kindergarten year. The parents participated as tutors. Three ways to implement and assess treatment integrity were used. To promote treatment integrity, the researchers organized an instruction meeting for all parents before the programme started. The first manual was presented and discussed. Exercises were demonstrated and discussed. After 1 week of training, the parents were phone-called to give them



the opportunity to ask questions and to find out if there was any lack of clarity. To check treatment integrity, the first author organized a second meeting after about 6 weeks of training. This gave us the possibility to investigate the application and solve possible problems. During this meeting the second manual was presented. To assess treatment integrity, two measures were used. The *quality* of the administration of the tutor programme was investigated by an assistant who visited the participants at home. During this visit, the assistant videotaped the execution of one tutor lesson. Afterwards the assistant and the parent discussed the lesson shortly. The visit lasted about 45–60 min and took place between the fifth and the tenth week of the programme. The videos were judged with the use of an observation instrument. The *quantity* aspect was measured by asking the participating parents to fill in daily log forms to provide information about the administration of the tutor programme. We asked the parents to send in the forms every fortnight by e-mail or traditional post-service. Participants who did not send in all forms by the time the children were post-tested, were contacted to find out how many lessons they had finished.

The children were tested individually with several pre-reading measures at their schools in a separate room before and after the training. Receptive vocabulary was measured before the intervention started and Nonverbal IQ was measured later, in first grade.

## Intervention

The programme, Sounding Sounds and Jolly Letters (Klinkende Klanken en Lollige Letters), was a home-based pre-reading programme with parents as tutors. It was a translation and adaptation of a Danish school-based kindergarten phoneme awareness and letter programme, 'Towards initial reading: Phonological awareness' (Borstrøm & Petersen, 1996). The Danish programme was described and applied in a longitudinal study (Elbro & Petersen, 2004). An addition we made to the training is the use of a mirror along with the articulation exercises and, more importantly, writing exercises of the newly learned letters on a slate. After two earlier studies with the programme, we extended the programme and made adaptations as a consequence of suggestions and remarks we got from the participating parents.

The Dutch programme was designed to take about 10 min a day, 5 days a week for 14 weeks (totally approximately 12 h). The programme focused on the letter–sound correspondences and gave special attention to the articulation of the speech sounds. It progressed slowly with approximately two speech sounds a week. The following 20 letters and speech sounds (12 consonants and eight vowels) were used in the programme: m, s, p, l, t, r, v, b, n, k, d, z, aa, oo, ee, uu, a, o, e, i (/m/, /s/, /p/, /l/, /t/, /r/, /v/, /b/, /n/, /k/, /d/, /z/, /a/, /o/, /e/, /y/, /a/, /ɔ/, /ɛ/, /I/). We used both long vowels (in Dutch represented by a digraph) and short vowels. Every speech sound and the corresponding letter were introduced in several ways. First, a rhyme or song was read to the child with focus on the speech sound. Then the child was shown a picture of the letter and was told the corresponding speech sound. If possible the child also received a semantic cue (/m/ is the 'taste-good' sound). After that the parent and child wrote down the letter several times at the slate. Together they thought of people

they knew whose name started with the sound, followed by an articulation exercise with attention to rounding (vowels), place of articulation and manner of articulation. The newly learned speech sounds were repeated in nursery rimes and language games. Phoneme blending and sound identity of both initial and final sound were trained by several language games. These games included the use of pictures, card games, and a hand doll. The other training materials consisted of two manuals (part one and part two), corresponding instruction sheets, a slate with chalk, and a progress table with funny stickers for the child. The training took place in the month of February–May 2003.

## Measures

### *Treatment integrity*

Apart from promoting and checking treatment integrity (see Procedure), we also assessed two aspects of treatment integrity: the quantity and the quality of the administration of the tutor programme.

*Quantity of administration.* We measured the quantity of administration, the number of weeks the subjects worked on the programme, with log forms for the parents to be filled in every day. On these forms they checked which lesson-parts were done and wrote down which problems were met. By contacting the parents when they did not return the log forms in due time, we made sure that the information was gathered.

*Quality of administration.* To measure the quality of administration we decided to videotape one lesson and to focus on the interaction between tutor and tutee. In particular we investigated whether the parent was able to give sufficient support at the right level and the child responded actively and enthusiastically. We did not restrict the operationalization of the quality of treatment integrity to the question whether the parent literally followed the manual text. Although the rhymes and songs and the other exercises were specified in the manuals and had to be presented, the programme allowed for some individual variation in the way it was carried out. More importantly, as is the case with many, if not all, programmes at this age, the important issue is not only whether the manual is followed, but also whether the quality of the interaction is reflected by involvement of the child.

The videos were analysed with the use of an observation instrument with five 5-point Likert scales. This instrument was a translation and adaptation of four Erickson-scales (Erickson, Sroufe & Egeland, 1985) and one self-made scale (marked below with an asterisk). The original Erickson scales were 7-point scales, designed to measure the quality of interactive behaviours of parent and child. We transformed them into 5-point scales to get a more global instrument, which was easier to supplement and adapt and therefore more suitable for our situation. Two raters were involved. We calculated the inter-rater-reliability (Weighted Kappa) over 25% of the videotaped lessons (eight lessons). The following scales were selected (inter-rater-reliability is put between parentheses): *Child persistence* ( $\kappa = 0.61$ ) measures to what degree the child was involved in the lesson and motivated to continue and complete the lesson; *Child enthusiasm* ( $\kappa = 1.00$ ) measures to what degree the child received the lesson with joy and enthusiasm; *Child responsiveness to parent* ( $\kappa = 0.89$ ) measures to what degree the child is willing to listen to and follow the instruction of the parent and accepts the

parents' authority; *Parent supportive presence* ( $\kappa = 0.90$ ) measures to what degree the parent shows emotional support to and confidence in the child; *Parent instruction adapted to the child\** ( $\kappa = 0.91$ ) measures to what degree the parent gives a clear instruction and adapts the instruction adequately to the level and needs of the child. We added this scale because we think that one of the most important aspects of the quality of interventions is the fit to the child. If the parent is able to adapt the instruction to the child well, this will probably lead to a persistent and enthusiastic child that learns a lot.

#### *Control measures and background variables*

*Receptive vocabulary.* This test is part of the Taaltoets Alle Kinderen (Language test All Children) (Verhoeven & Vermeer, 2001). Each item consisted of four pictures and a spoken word. The child has to choose the alternative that best matches the given word. The test had 98 items that increased in difficulty. The administration of the test was stopped when the child failed six or more of the last eight items. For this test, the Cronbach's alpha was 0.89.

*Nonverbal IQ.* We used the Coloured Progressive Matrices to measure nonverbal IQ (Raven, Court, & Raven, 1984). The test has three subtests with 12 test items each. Every item consists of a rectangular pattern in which a part is missing. The child has to look for the missing part and chose between six alternatives. The maximum score is 36. Split-half-reliability (corrected for test-reduction) for 6-year olds was 0.82 and for 7-year olds 0.84 (van Bon, 1986).

*Age of the child, education level of the father, and education level of the mother* were measured using the 'Vragenlijst Vroege Diagnose en Behandeling van Ontwikkelingsdyslexie' (Questionnaire Early Diagnosis and Treatment of Developmental Dyslexia) (see Eleveld, 2005). The completed education level of both the father and the mother of the child were measured on a 7-point ordinal scale, ranging from 1 (*elementary school*) to 7 (*university degree*).

#### *Pre-reading skills*

*Receptive letter knowledge (pre- and post-test).* Receptive knowledge of 32 letters or digraphs was tested. Letters with a very low frequency in Dutch (e.g. c, q, x, y) were not presented. The experimenter showed the child a sheet with rows with six printed lowercase letters each. Then she gave a speech sound and asked the child to indicate the letter in that row that matched the sound. Every correct indicated letter or digraph gave one point (Verhoeven, 2002). Reported reliability (Cronbach's alpha) in second kindergarten year was 0.87–0.89 (Eleveld, 2005).

*Productive letter knowledge (post-test only).* For this purpose we used the 'Grafementoets' (Grapheme test) (Verhoeven, 1993a). The child has to read out loud 34 separate graphemes. The maximum score was 34. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

*First-sound identity (pre- and post-test).* The experimenter required the child to choose from four alternatives the word that had the same first sound as a given word (e.g. *saw* begins with /s/). Each item consisted of a row of five pictures. The first picture represented the given word and was separated from the other pictures by a vertical line. The experimenter told the names of the pictures and also the first sound of the given word. The child was instructed to identify the picture of the word with the same sound at the beginning (e.g. *tree, sun, ball*, or

*comb*). The test consisted of two practice items and 10 test items and had a maximum score of 10. Reported reliability for this test was 0.86–0.87 (Eleveld, 2005).

*Phoneme blending (pre- and post-test)*. Toets voor Auditieve Synthese (Test for Phoneme Blending) (Verhoeven, 1993a). The experimenter gave the child a word in separate phonemes and asked the child to blend the phonemes and pronounce the correct word (e.g. /p/-/l/-/a/-/n/-/t/ makes *plant*). The maximum score was 20. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

*Phoneme segmentation (pre- and post-test)*. Toets voor Auditieve Analyse (Test for Phoneme Segmentation) (Verhoeven, 1993b). The experimenter gave a whole word and the child was required to segment the word in separate phonemes (e.g. *raam* [window] makes /r/-/a/-/m/). The maximum score was 20. Cronbach's alpha was above 0.85 (Verhoeven, 2000).

*Decoding (post-test only)*. Woordbouw Leestaak (Word Building Reading Task) (Regtvoort, 2003). The experimenter asked the child to read out loud 13 CVC-words and five CVCC-words. Every well-read word gave 1 point. The maximum score was 18. No reliability data were available of this experimental task.

## RESULTS

### Level of and Variation in Treatment Integrity

#### *Quantity of administration*

The histogram in Figure 1 shows that most families (66%) completed the whole training programme (14 weeks). A considerable proportion of families (34%), however, did not finish the complete programme. Some stopped early in the programme and some later on, mainly because of lack of time. The mean number of weeks the families completed was 11.91 (*S.D.* = 3.55). Because the variable quantity of administration was extremely skewed (skewness =  $-1.57$ ), we made this variable dichotomous for further analysis. Because the median turned out to

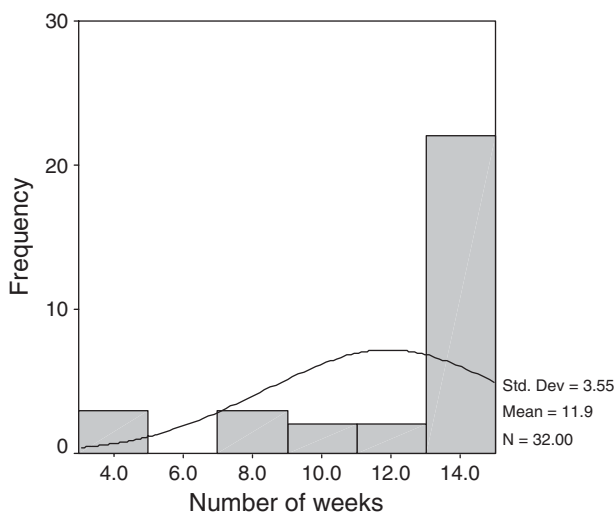


Figure 1. Frequencies of the quantity of administration.

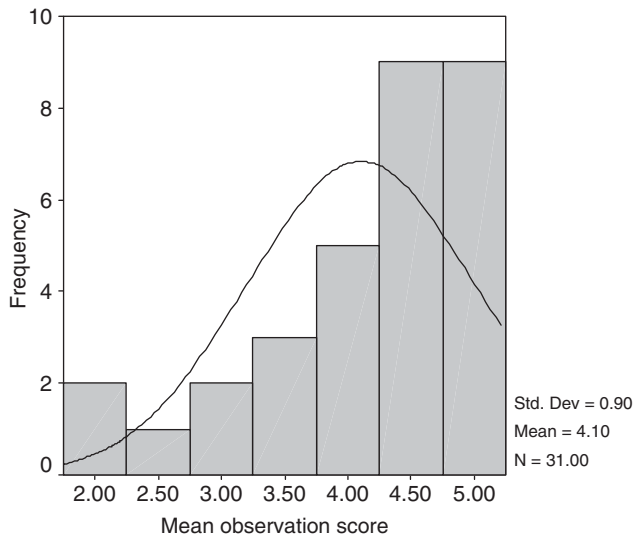


Figure 2. Frequencies of the quality of administration.

be 14 weeks and we wanted the numbers in each category to be comparable, we made the following two categories: 1 (completed whole programme) and 0 (did not complete whole programme).

#### *Quality of administration*

A factor analysis showed that the five 5-point observation scales measuring the quality of administration loaded on one factor, indicating the interdependency of parental activation and involvement of the child. Factor loadings of the principal components analysis were between 0.86 and 0.93, and 78% of variance was explained. Therefore, we decided to make one variable of the mean observation score, which we called *quality of administration*. The Cronbach's alpha was 0.92. The histogram in Figure 2 shows that a large number of families (74%) got a high mean observation score, a score between 4 and 5. From one family, the observation scores were missing. The missing values were therefore estimated and imputed by a missing value analysis using the regression-method (SPSS 10.0 Syntax Reference Guide, 1999). The variable quality of administration was skewed (skewness =  $-1.19$ ) and had a mean score of 4.10 ( $S.D. = 0.89$ ). In some families, the dyslexic parent worked with the children on the programme, but in most cases the other parent worked with the children. In most cases only the mother worked with the programme (59%) and in most cases only the father was dyslexic (77%). The dyslexic participants that worked with the programme reported no problems, but some needed an extensive preparation for each lesson.

### **Prediction of Pre-Reading Skills at Post-Test**

#### *Pre- and post-test*

First we transformed the raw scores on all pre-reading skills measures into the proportion of correct answers. After this, we did a factor analysis on the four

pre-reading skills pre-test scores. All pre-tests, receptive letter knowledge, first-sound identity, phoneme blending, and phoneme segmentation, loaded on one factor. Factor loadings of the principal components analysis were between 0.81 and 0.95, and 77% of the variance was explained. We made a new variable by calculating the mean of these four pre-test variables and called it *pre-reading skills pre-test*. The Cronbach's alpha was 0.89. The new variable was somewhat skewed and substantially flattened (skewness = 0.84, kurtosis = -2.29) and had a mean score of 37.91 (*S.D.* = 23.89). The same procedure was followed for all six pre-reading post-tests, same tests as pre-tests completed with productive letter knowledge and decoding. It appeared that all dependent variables also loaded on one factor. Factor loadings of the principal components analysis were between 0.73 and 0.90, and 72% of the variance was explained. Again we calculated a new variable, which we called *pre-reading skills post-test*. The Cronbach's alpha was 0.90. The new dependent variable was normally distributed and had a mean score of 50.34 (*S.D.* = 21.05).

### *Intercorrelations*

Before we executed regression analyses to investigate whether the level of pre-reading skills at the end of kindergarten could be predicted by treatment integrity of the tutor programme, a Pearson correlation matrix was calculated between the predictors, the dependent variable and the available control and background measures (see Table 3). Because two background measures, education level of both the father and the mother, were measured ordinal on a 7-point scale, and regression analyses require higher levels of measurement (interval or ratio), we made these variables dichotomous. The two categories were: 1 (university or college degree, score 6 or 7) and 0 (other, score 1–5).

The results of the correlation matrix showed that education level of the mother and educational level of the father were intercorrelated fairly well ( $r = 0.58$ ). Pre-reading skills pre-test was positively correlated with the control measures receptive vocabulary, education level of the mother and educational level of the father. No other control variables were significantly intercorrelated. Not surprisingly, the dependent variable, pre-reading skills post-test, was highly correlated with pre-reading skills pre-test ( $r = 0.86$ ). The dependent variable also correlated fairly well with the predictors quantity of administration ( $r = 0.60$ ) and quality of administration ( $r = 0.45$ ). The dependent variable did not correlate with either age of the child or nonverbal IQ of the child. Therefore we excluded these variables from the regression analyses.

### *Regression analyses*

Because of the rather small sample size ( $N = 32$ ), we could not include too many independent variables in a regression analysis. For a reliable regression equation it has been proposed that about 15 subjects per independent variable are needed (Stevens, 1996, p. 125). We decided to omit education level of the father for three reasons: (1) it correlated fairly well with another control measure, education level of the mother, (2) the correlation with the dependent variable was smaller than of other control variables, and (3) it was not correlated with the predictors, whereas the education level of the mother was. Also the fact that in most families (59%) the mother was the tutor influenced this choice. In 16% of the families the father

Table 3. Intercorrelations between predictors, background variables, child variables, and the dependent variable (*N* = 32)

	Variables								
	1	2	3	4	5	6	7	8	9
Predictor	—	0.34 <sup>†</sup>	0.34 <sup>†</sup>	0.20	0.20	0.03	0.02	0.33 <sup>†</sup>	0.45 <sup>**</sup>
Predictor		—	0.11	−0.03	0.33 <sup>†</sup>	0.10	−0.13	0.31 <sup>†</sup>	0.60 <sup>**</sup>
Covariate			—	0.58 <sup>**</sup>	−0.01	−0.05	0.27	0.49 <sup>**</sup>	0.45 <sup>**</sup>
Covariate				—	−0.27	−0.10	0.25	0.39 <sup>*</sup>	0.30 <sup>†</sup>
Covariate					—	0.19	0.10	−0.04	0.15
Covariate						—	−0.02	0.42 <sup>*</sup>	0.37 <sup>*</sup>
Covariate							—	0.23	0.24 <sup>**</sup>
Dependent								—	0.86 <sup>**</sup>

Note: dichot., dichotomous.  
<sup>†</sup>Correlation is significant at the 0.10 level (2-tailed).  
<sup>\*</sup>Correlation is significant at the 0.05 level (2-tailed).  
<sup>\*\*</sup>Correlation is significant at the 0.01 level (2-tailed).

Table 4. Separate hierarchical regression analyses for treatment integrity variables predicting pre-reading skills at post-test after controlling for different (combinations of) background- or child variables ( $N = 32$ )

Analysis	Step	Variable	Pre-reading skills post-test		
			$\beta$	$\Delta R^2$	$R^2$ total
1	1	Quantity of administration	0.51**	0.36***	0.43***
	2	Quality of administration	0.28 <sup>†</sup>	0.07 <sup>†</sup>	
2	1	Receptive vocabulary	0.31*	0.13*	0.53***
	2	Quantity of administration	0.47**	0.32***	
	3	Quality of administration	0.29*	0.07*	
3	1	Education level of the mother	0.34*	0.21**	0.54***
	2	Quantity of administration	0.51***	0.31***	
	3	Quality of administration	0.17	0.02	
4	1	Pre-reading skills pre-test	0.72***	0.74***	0.87***
	2	Quantity of administration	0.34***	0.12***	
	3	Quality of administration	0.11	0.01	
5	1	Pre-reading skills pre-test	0.74***	0.74***	0.86***
	2	Receptive vocabulary	0.02	0.00	
	3	Quantity of administration	0.37***	0.12***	
6	1	Pre-reading skills pre-test	0.71***	0.74***	0.86***
	2	Education level of the mother	0.06	0.00	
	3	Quantity of administration	0.37***	0.12***	

<sup>†</sup> $p = 0.07$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

and in 25% of the families both the father and the mother or other family members worked with the programme.

Separate hierarchical regression analyses were conducted. The results are presented in Table 4. The first analysis indicated that pre-reading skills at post-test could be predicted by treatment integrity. Quantity and quality of administration accounted for 43% of the variance in pre-reading skills at the post-test. The second analysis showed that the two treatment integrity measures still contributed to the prediction of pre-reading skills post-test after controlling for receptive vocabulary. The total regression model accounted for 53% of the variance. The third analysis showed that after controlling for education level of the mother, quantity of administration still contributed to the prediction of pre-reading skills post-test. Quality of administration, however, did not significantly contribute to the prediction any more. The proportion explained variance was 54%. In the fourth analysis, we controlled for pre-reading at pre-test. This time, the proportion explained variance was much larger (87%). Although the contribution of pre-reading skills pre-test to the prediction equation was very large (74% of variance), and it eliminated the contribution of receptive vocabulary and education level of the mother (see analyses 5 and 6), the quantity of administration still contributed significantly to the prediction equation.

We also analysed if there were any two-way interactions. We found a significant interaction of education level of the mother  $\times$  quality of administration,  $\Delta R^2 = 0.18$ ,  $F(1,28) = 9.72$ ,  $p < 0.01$ , and a tendency to a significant



interaction of education level of the mother  $\times$  quantity of administration,  $\Delta R^2 = 0.06$ ,  $F(1,28) = 4.15$ ,  $p = 0.051$ . To interpret these interactions we did separate analyses for the group with high-educated mothers ( $n = 17$ ) and for the group with low-educated mothers ( $n = 15$ ). In the group with high-educated mothers the contribution of the quality of administration to the prediction of the post-test was considerable ( $R^2 = 0.49$ ), whereas in the group with low-educated mothers quality did not contribute to the prediction ( $R^2 = 0.00$ ). The quantity of administration contributed to the prediction in families with high-educated mothers as well as in families with low-educated mothers. However, the contribution to the prediction was much larger in families with high-educated mothers ( $R^2 = 0.56$ ) than in families with low-educated mothers ( $R^2 = 0.22$ ). Furthermore, a  $t$ -test showed that children with high-educated mothers outperformed the children with low-educated mothers on both the pre-test,  $t(30) = 3.08$ ,  $p < 0.01$ ,  $d = 1.13$ , and the post-test,  $t(30) = 2.79$ ,  $p < 0.01$ ,  $d = 1.02$ .

## DISCUSSION

In the current study, treatment integrity was promoted, checked as well as assessed. As expected, most parents finished the whole intervention programme. However, about one-third stopped before finishing the programme, mainly because lack of time. The quality of the administration of the programme was high (observation score of 4 or above) in most cases. Transforming the observation score into a percentage and taking full completion of the programme as the quantitative criterion, we may conclude that the assessed treatment integrity was on average about 74% for quality and 66% for quantity.

Because there turned out to be variation in the two measures of treatment integrity, we were able to investigate the influence of treatment integrity on the dependent variable. The results showed that, in line with our expectations, treatment integrity contributed to the prediction of pre-reading skills at the end of kindergarten. These findings support the idea that treatment integrity is an important topic in intervention research.

The contribution to the prediction equation of the quantity of administration turned out to be considerable. Without controlling for other variables, the quantity of administration accounted for 36% of the variance in the dependent variable. After controlling for receptive vocabulary of the child or the education level of the mother, the quantity-measure still accounted for more than 30% of the variance. After controlling for pre-reading skills at pre-test, the contribution of the quantity of administration was reduced considerably, but remained statistically significant (explaining 12% of the variance).

The contribution of the quality of administration to the prediction of pre-reading skills at the end of kindergarten was smaller. After controlling for the quantity of administration, this measure accounted for 10% of the variance in the dependent variable. After controlling for both quantity of administration and the education level of the mother or pre-reading skills at pre-test the contribution vanished. Although this is partly caused by the order of the treatment integrity variables, the contribution of quality remained smaller than that of quantity, when they were entered in reverse order.

A possible explanation for the small contribution of the quality of administration to the prediction equation can be that it was assessed only once. Using videotaped observations of one session only, we assumed that the quality of interaction of parent and child was stable over time (*time generalizability*; see Gresham *et al.*, 2000). However, the parent and the child may have shown higher quality of interaction when the observer was present (*observer reactivity*; see Gresham *et al.* (2000)). In general, one observation may not be sufficiently representative for a series of sessions to get a reliable picture. Another explanation for the relatively small contribution of quality may be that a certain threshold value represents sufficient quality and quality scores above the threshold do not result in better scores on the dependent variable. If this is the case, the variation in quality scores was probably insufficient to show larger results. After all, Figure 2 showed that the distribution is skewed and relatively high scores are over represented.

The finding that the quantity of administration was an important predictor of treatment outcome supports the internal validity of the intervention programme because finishing it completely made it more effective. Furthermore, the fact that quantity was a better predictor than quality of administration is reassuring for two reasons. Although parents are nonprofessionals with regard to these type of interventions, it seems to be more important that they finish the programme than that they interact with the child in a superior way. Moreover, the quantitative aspect of treatment integrity is easier to implement and measure and therefore more cost-effective. When we consider the methodological importance of treatment integrity and we may assume that it can be assessed so easily, it is even more surprising that quantitative aspects of treatment integrity are neglected in much of the phoneme awareness intervention research.

The role of treatment integrity varied with the education level of the mother. In families with a low-educated mother, the quality of the interaction between mother and child did not contribute to the prediction of pre-reading skills at the end of kindergarten, while in families with a high-educated mother it did so considerably. However, it should be noted that the relevance of this finding is limited, as the quality of administration did not contribute to the prediction of pre-reading skills at the end of kindergarten after pre-reading skill at pre-test was controlled for. The number of lessons that were completed seemed also more important in families with a high-educated mother, although the difference between the groups was less strong.

Our data indicated that children with high-educated mothers had on average better pre-reading skills than children with low-educated mothers (both at pre- and post-test). This finding supports the suggestion that a home environment that is linguistically rich positively influences early literacy skills (see for instance Jordan, Snow, & Porche, 2000). Higher educated mothers probably created a more literate environment for the child. The correlation between the education level of the mother and the pre-reading skills of the child indeed was considerable in our study. However, it should be noted that the education level of the mother is not the only explanation for higher reading skill. After all, the education level of the mother (a distal variable) interacted with treatment integrity (a more proximal variable). More importantly, participation in the intervention programme seems to provide some escape from this mechanism. Table 3 showed that the role of the education level of the mother slightly decreases in time (pre-test to

post-test) while the role of quality and quantity of administration slightly increases in time.

It may be argued that combining child scales and parent scales is unusual or even inappropriate in measuring the quality of administration of the intervention. But in line with the transactional approach of development (see for instance Sameroff & Chandler, 1975; Sameroff & MacKenzie, 2003), one can interpret both parent variables and child variables as part of the interaction between them. After all, in the transactional model the development of any process in the individual (the child) is influenced by interaction with the individual's context (the parent). Besides, the factor analysis showed that all observation scales loaded on one factor, meaning they mainly measure the same construct, most likely a transactional construct.

It also may be argued that the quality of the administration, in the way we measured it, is not only an aspect of treatment integrity, but also an aspect of the pedagogical climate of the home environment of the child. It includes the didactical skills of the parent and the responsiveness of the child. It may be that part of the quality of the parent-child interaction is specific for this phoneme awareness and letter-knowledge intervention and part of the quality of the parent-child interaction is a proximal expression of a more distal aspect, the pedagogical climate. In future research, using a larger sample, it may be worthwhile to control for distal influences by measuring the parent-child interaction with different tasks apart from the intervention.

In many studies the direction of causality remains a problem. In our study it remains uncertain whether treatment integrity influences the pre-reading skills of the child or vice versa. It is conceivable that families with children who easily learn letters and phoneme awareness skills will show high quality of interaction in working on the programme and complete many lessons. We tried to deal with this causality issue by controlling for the starting level of the children in pre-reading skills and their verbal intelligence (receptive vocabulary). After controlling for both, the quantity of administration still contributed to the prediction. Besides, both treatment integrity measures correlated less with pre-reading skills at pre-test than at post-test. This suggests that treatment integrity is less influenced by pre-reading skills at the pre-test than treatment integrity influenced these skills at the post-test.

Although the findings of our study clearly underline the importance of assessment of treatment integrity, the size of our sample ( $N = 32$ ) was small in order to allow prediction analyses. We suggest that replications of intervention studies also should focus on treatment integrity and should include larger subject samples.

It should be noted that treatment integrity could never be a predictor of outcome on its own. Apart from Hawthorne effects, when an intervention is ineffective but is implemented correctly and faithfully, it is still ineffective. The treatment integrity in this study only contributes to the prediction of pre-reading at the end of kindergarten because the intervention was effective. However, whether the outcome of an intervention study can be ascribed to the intervention remains unclear when the integrity of treatment is not known. Because we assessed the level of treatment integrity and it turned out to be sufficient, we may conclude with more confidence that the outcome of the intervention was most likely caused by the programme and its implementation.

A practical question related to this study is, is it prudent to let parents of at-risk children work autonomously with such an intervention. Because training of phoneme awareness and letter knowledge does not belong to the pedagogical means of the average parent, we suggest that professional guidance is obligatory.

Although in our study treatment integrity was promoted and checked as well as assessed, parent and child dyads still differed in the quantitative and qualitative aspects of treatment integrity, which turned out to be related to treatment outcome. This result also indicates that implementation should be adapted more to the needs of parents, for example, to give parents who are relatively less educated more support in order to enhance the quality of the home literacy environment of the children.

## ACKNOWLEDGEMENTS

We thank the participating children, parents and schools for their time and effort. We also thank the research assistants including Anna Pires and Sharon Geurtsen for their help in collecting the data.

## REFERENCES

- Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology*, 86(2), 235–243.
- Barrs, M., & Ellis, S. (1998). *Reading together parents' handbook*. London: Walker Books.
- Borstrøm, I., & Elbro, C. (1997). Prevention of dyslexia in kindergarten: Effects of phoneme awareness training with children of dyslexic parents. In C. Hulme, & M. Snowling (Eds.), *Dyslexia: Biology, cognition and intervention* (pp. 235–253). London: Whurr.
- Borstrøm, I., & Petersen, D. K. (1996). *På vej til den første læsning. Fonologisk opmærksomhed* [On the way to initial reading. Phonological awareness]. Copenhagen: Alinea.
- Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training. An intervention study with children at risk of dyslexia. *Journal of Educational Psychology*, 96, 660–670.
- Eleveld, M. A. (2005). *At risk for dyslexia. The role of phonological abilities, letter knowledge, and speed of serial naming in early intervention and diagnosis*. Ph.D. thesis, State University Groningen, The Netherlands. Leuven: Garant Publishers.
- Erickson, M. F., Sroufe, L. A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. *Monographs of the Society for Research in Child Development*, 50, 147–166.
- Fielding-Barnsley, R., & Purdie, N. (2003). Early intervention in the home for children at risk of reading failure. *Support for Learning*, 18(2), 77–82.
- Gresham, F. M., Gansle, K. A., Noell, G., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*, 22(2), 254–272.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, 15(4), 198–205.
- Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65(1), 41–57.

- Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. P. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology*, 97(4), 687–704.
- Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly*, 35(4), 524–547.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment integrity in outcome studies. *Clinical Psychology Review*, 11(3), 247–266.
- O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. A. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children*, 59(6), 532–546.
- O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology*, 87(2), 202–217.
- Petersen, D. K. (2000). What do you get if you add /mmm/ to ice? Training phoneme awareness in kindergarten: An intervention study with children of dyslexic parents. In N. A. Badian (Ed.), *Prediction and prevention of reading failure* (pp. 247–273). Baltimore: York Press.
- Rack, J. (2004). The theory and practice of specialist literacy teaching. In G. Reid, & A. Fawcett (Eds.), *Dyslexia in context. Research, policy and practice* (pp. 120–131). London and Philadelphia: Whurr Publishers.
- Raven, J. C., Court, J. H., & Raven, J. (1984). *Manual for Raven's progressive matrices and vocabulary scales*. London: Lewis.
- Regtvoort, A. G. F. M. (2003). *Woordbouw leestaak [Wordbuilding reading task]*. Amsterdam: University of Amsterdam.
- Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Siegal (Eds.), *Review of child development research* (Vol. 4, pp. 187–244). Chicago: University of Chicago Press.
- Sameroff, A. J., & MacKenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology*, 15, 613–640.
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- Slocum, T. A., O'Connor, R. E., & Jenkins, J. R. (1993). Transfer among phonological manipulation skills. *Journal of Educational Psychology*, 85(4), 618–630.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly*, 34(1), 28–52.
- van Bon, W. H. J. (1986). *Raven's coloured progressive matrices; Nederlandse normen en enige andere uitkomsten van onderzoek [Dutch standards and some other outcomes of research]*. Lisse, the Netherlands: Swets & Zeitlinger.
- Verhoeven, L. (1993a). *Grafemetoets. Toets voor Auditieve Synthese*. Handleiding. [Grapheme test. Test for phoneme blending. Manual]. Arnhem, The Netherlands: Cito.
- Verhoeven, L. (1993b). *Toets voor Auditieve Analyse*. Handleiding [Test for phoneme segmentation. Manual]. Arnhem, The Netherlands: Cito.
- Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading*, 4(4), 313–330.
- Verhoeven, L. (2002). *Passieve Letterkennis [Receptive letter knowledge]*. Arnhem, The Netherlands: Cito.
- Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen—Handleiding [Language test All Children Manual]*. Arnhem, The Netherlands: Cito.

- Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, 30(5), 679–689.
- Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology*, 72, 1–15.